20th International Scientific Conference AMSE
Applications of Mathematics and Statistics in Economics 2017
Szklarska Poręba, Poland 30 August 2017 – 3 September 2017

METHODS FOR THE ESTIMATES OF REGIONAL ECONOMIC STRUCTURES

KAREL ŠAFR, JAROSLAV SIXTA
University of Economics in Prague, Department of Economic Statistics,
Nám. W. Churchilla 4, Prague 3, Czech Republic
email: karel.safr@vse.cz, sixta@vse.cz

Abstract

The paper brings an overview of statistical methods used for estimation of regional economic structures. The methods are presented using the case of the Czech Republic and Czech regional Input-Output tables for 2013. Several models used and derived for these purposes are presented within the paper, and their results are subjected to sensitivity analysis with the aid of the CGE model. Regional economic structures are necessary for advanced economic analyses aimed at the regional level, and since they are not published by official statistical agencies, they must be based on academic research. The core of regional structures is derived from both national and regional Input-Output tables and, therefore, they provide a good basis for the use of Input-Output analysis. From the statistical point of view, computational process results in a big square matrix depicting the flows of products allowing the use of dynamic Input-Output models. There can be found several methods for obtaining the flows between the regions and we briely focus on gravity method, optimalization of distance, physical flows and distance betweem the centroids. The results presented within the paper should be useful for economists dealing with regional economic models.

Key words: *regional Input-Output, interregional flows, computable general equilibrium*

JEL Codes: *C68, D57, R13*

DOI: 10.15611/amse.2017.20.33

1. Introduction

Input-Output tables represent a very detailed tool for describing the economy. Their use is wide-ranging, from a descriptive-statistical tool to the analytical grasping of this data source using structural macroeconomic models. The regional aspect of national accounts then adds a further level of detail to this tool. This level of detail is particularly important in the context of non-homogeneous regional structures – i.e. for countries whose regions have a different economic structure. From the perspective of describing the economy, heterogeneity of regions is not a significant problem – assuming for the entire economy. From an analytical point of view, however, heterogeneity can cause significant distortions in the effects being evaluated, especially if impacts occur in one region.

Regional Input-Output tables as such limit the analysis to the regions as entities without the context of links to other regions. Therefore, another important element is the interregional flows matrix which is used in the multiregional analysis. In terms of the Czech Republic, the regional I-O tables have been compiled for 2011 and 2013. However, these matrices do not include interregional production flows that can be estimated using several approaches. We are mainly concerned with modelling-based methods and hybrid methods based on indirect data sources. In this paper, we compare the gravity model, import distance minimization model, proportional distribution model, inverse distance matrix model and model using the method of approximation based on physical volumes of production flows. Both the gravity model method and the estimation method based on physical volumes have the same underlying data which only differs in the approach to distance estimation.

The aim of this paper is to present the construction of a standard regional model of equilibrium, i.e. the CGE model, and to illustrate the sensitivity analysis of interregional flow estimations based on regional Input-Output tables.

2. Literature

With regard to the popularity of Input-Output analysis, a very wide range of applications of this tool to data sources can be found. For example, a model of multiregional table combination or the CGE model for modelling the impact of motorway construction (Kim, Hewings et al., 2004) can be mentioned in the context of regional economies.

The article by Sargento et al. (2012) can be considered a comprehensive study summarizing the basic methods of regional flow estimation. The authors summarize the basic methods of interregional flow estimation here. In this context, it is also important to mention the user manual issued by ESCAP and ARTNet (Sheperd, 2013), which summarizes a detailed approach and methods of estimating gravity models. In terms of the Czech Republic data, interregional flows were estimated in the final thesis by Kieslichová (2016), but the author did not use the possibility of calibrating the model for export and import flows. Furthermore, in Šafr (2016a), in both previous articles, the gravity method was estimated using an approach without knowledge of flows and, therefore, elasticity estimation, which, as shown in this paper, leads to the unintended simplification of NGM in the RAS method[1] based on inverse distances between regions.

In terms of data calculations in general (impact of non-survey updates), Input-Output tables and SAM matrices year-on-year were dealt with by Cardenete and Sancho (2004). In this study, they dealt with changes in Input-Output data and their effect in the regional CGE model. Using a simulated change in the tax rate, they concluded that year-on-year changes do not have a significant impact on the CGE model results. However, the authors do not assess the impact of the effect of the choice of different data estimation methods and primarily deal with RAS optimization. Similarly, Šafr (2017) compared the impact of different methods of estimating capital matrices (Šafr, 2016b) on simulated shocks in the IO and CGE models. With regard to capital formation matrices, it was shown that in the case of the I-O model the differences between the approaches were more noticeable in the results of the model. However, in the case of the CGE model, the differences in impacts on the economy are negligible.

3. Production flow matrix

The key element of the multiregional model is the production flow matrix which describes the use of products in regional distribution and distinguishes between the export and import regions. The rows of such a matrix show the regional submatrices of the regions exporting goods, and the columns show the regions importing goods, whether for intermediate consumption or end use. Since the entries of these matrices are unknown, the purpose of the individual methods is to estimate them. However, we know some partial information about this matrix: which regions do not export and import goods at all (reduction of task dimension),

[1] The RAS method is a general method for I-O balancing (or any matrix). The aim is to achieve a new I-O data structure from a predefined (prior) structure. In this task are known the new row and column totals. We are noting this method as a function f_{RAS}.

as well as the row sums or, as the case may be, column sums of this matrix. This means that we can define this matrix as matrix X^{TOT}, for which the following applies:

$$\mathbf{X}^{TOT} = \begin{bmatrix} \mathbf{X}^{11} & \mathbf{X}^{1r} & \mathbf{X}^{1R} \\ \mathbf{X}^{p1} & \mathbf{X}^{pr} & \mathbf{X}^{pR} \\ \mathbf{X}^{P1} & \mathbf{X}^{\mathrm{Pr}} & \mathbf{X}^{PR} \end{bmatrix}, \text{ where } \mathbf{X}^{pr} = (x_{ij}^{rp})_{nn}, \text{ for } r, p = 1,2......P, (P{=}R) \quad (1)$$

and, therefore, the size of $\mathbf{X}^{TOT}$ is $(Rn) \times (Rn)$.

The diagonal submatrices of matrix $\mathbf{X}^{TOT}$ (those where $r = p$) represent a regional production flow and, in addition to diagonal submatrices, they represent the flow of production from region *p* to region *r*. Here, it is important to point out that diagonal submatrices are generally considered to be known in this task and represent a single-regional model.

Aggregate production flows between regions can be expressed as follows:

$$x_{\bullet\bullet}^{pr} = \sum_{i}^{n}\sum_{j}^{n} x_{ij}^{pr}, \text{ for } r, p = 1,2......P, (P{=}R), \text{ and } i, j = 1,2......n. \quad (2)$$

Where $x_{\bullet\bullet}^{pr}$ represents the aggregate flow of production between region *p* and region *r*. This flow is the total value of exports from region *p* to region *r* and, conversely, the total value of imports of region *r* from region *p*.

In addition to diagonal matrices, they can be arranged in "flow matrices" for individual products. The flow matrix then represents the following:

$$\mathbf{F}_i = \begin{bmatrix} 0 & x_{i\bullet}^{1r} & x_{i\bullet}^{1R} \\ x_{i\bullet}^{p1} & 0 & x_{i\bullet}^{pR} \\ x_{i\bullet}^{P1} & x_{i\bullet}^{\mathrm{Pr}} & 0 \end{bmatrix}, \text{ where } x_{i\bullet}^{pr} = \sum_{j}^{n} x_{ij}^{pr} \quad (3)$$

for $r, p = 1,2......P$, (P=R), and $i, j = 1,2......n$.

This matrix expresses production flows for industry *i* from region *p* to region *r*. x_i^{pr} then expresses the total imports of product *i* from region *p* to region *r*. With regard to this matrix, it is important that we know its row sums and column sums. Therefore, they are known:

$$x_{i\bullet}^{p\bullet} = \sum_{r}^{R} x_{i\bullet}^{pr} \quad \text{and} \quad x_{i\bullet}^{\bullet r} = \sum_{p}^{P} x_{i\bullet}^{pr}, \quad \text{for } r, p = 1,2......P, \ (P{=}R), \text{ and } i, j = 1,2......n. \quad (4)$$

Where $x_i^{p\bullet}$ represents exports of product *i* from region *p* to all other regions and, conversely, $x_i^{\bullet r}$ represents imports of product *i* in region *r* from all other regions.

Using estimation methods, we then assume certain characteristics of the entries of this matrix (F_i). In general, we assume that the diagonal entries of the matrix are zero (they would represent the region's own production, which is, however, known). We also know its column sums and row sums – which is a limitation of the models. The latest known information of this task is the value of total imports of other products in product *j* to region *r*, or the inverse sum for the preceding equation:

$$x_{\bullet j}^{\bullet r} = \sum_{p}^{P}\sum_{i}^{n} x_{ij}^{pr}, \text{ for } r, p = 1,2......P, (P{=}R), \text{ and } i, j = 1,2......n. \quad (5)$$

This information is then used by the following methods to estimate interregional production flows.

4. Data estimation methods

We can divide the methods for the calculation of regional production flows into three basic groups. These are: 1) data surveys, 2) model calculations based on related data sources, and 3) model calculations without data sources. Data surveys are considered as very demanding method. Researchers would have to investigate representative number of companies (n firms in R regions). Next, gravitational methods can serve as an example for the method of the second type. Finally, the method of distance optimization belongs as an example of the calculation method for the third group.

4.1 Gravity method (GM)

The gravity method is based on Newton's law of gravitation. This law states that the force of attraction between two objects is proportional to the product of their masses and inversely proportional to the square of the distance between them (and multiplied by the gravitational constant). The first use of this physical law in this task was designed to estimate production flows between towns in the Middle Ages. The method is usually based on an equation by which we can estimate the entries of matrix F_i:

$$\widetilde{\mathbf{F}}_i = (\widetilde{x}_{i\bullet}^{pr})_{r\times r}, \text{ where: } \widetilde{x}_{i\bullet}^{pr} = G_i \frac{(x_{i\bullet}^{p\bullet})^{\alpha_i}(x_{i\bullet}^{\bullet r})^{\beta_i}}{\left(\delta^{pr}\right)^{\omega_i}}, \tag{6}$$

for $r, p = 1,2......P$, (P=R), and $i, j = 1,2......n$.

Thus, the flow of production from region *r* to region *p* ($\widetilde{x}_{i\bullet}^{pr}$) is proportional to the force of their exports and imports[2] of product *i* ($x_{j\bullet}^{p\bullet}$ and $x_{j\bullet}^{\bullet r}$) and distance δ^{pr} between these two regions. It should also be noted that $\widetilde{x}_{i\bullet}^{pr}$ represents the initial estimate of the entries of matrix $\mathbf{F}_i$. Furthermore, it is necessary to calibrate this estimate of the coefficients ($\widetilde{x}_{i\bullet}^{pr}$) with respect to the inconsistencies of the estimated entries with the model conditions (equations 4 and 5). Constants α_i, β_i and ω_i can then be calibrated or estimated by regression. These constants are usually interpreted as distribution of production/distance. Constant G_i represents the gravitational constant of Newton's model but, in our case only, it only sets the level of production flows for product *i*. In order to estimate the model constants (α_i, β_i and ω_i), however, it is necessary to know the production flow ($x_{i\bullet}^{pr}$), which is subject to estimation – it is not known.

There are several ways to proceed at this point. Either we know the aggregate flows between the regions ($x_{\bullet\bullet}^{rp}$ – sum of all product flows that are exported from region *r* to region *p*). Then the aggregate flows of the model constants ($\alpha = \alpha_i, \beta = \beta_i$ and $\omega = \omega_i$) can be estimated. For a small number of regions, they can be calibrated. For many regions, regression can be applied, using logarithmic adjustment:

[2] Exports and imports are not usually used here, as export and import volumes are mostly unknown. However, we know the sum of total imports (from all regions) and the sum of total exports (to all other regions), so we used the equation in a modified form.

$$\log(x_{i\bullet}^{pr}) = \log(G_i) + \alpha \log(x_{i\bullet}^{p\bullet}) + \beta \log(x_{i\bullet}^{\bullet r}) - \omega \log(\delta^{pr}) \tag{7}$$

for $r, p = 1,2......P$, (P=R), and $i, j = 1,2......n$.

The observed data for the individual production flows can be used for such estimated coefficients. After estimating these matrices, it is still necessary to apply the nonlinear RAS method to ensure that the conditions for estimating the matrices of interregional production flows are met, see equations 4 and 5.

The second way is to estimate the gravity model based on estimations from the matrices of physical flow volumes between regions. If we know the matrices of production flows between regions in physical volumes, then we can estimate the coefficients using this structure. Subsequently, using the nonlinear RAS method, the resulting matrices of structures can be estimated for the products. In this study, we will especially use the second method, where we will calibrate the gravity model for individual products from the matrices describing production flows in the NST classification[3] (NST01 – NST20). The advantage of this approach is a more detailed estimate of the coefficients than when estimating coefficients in one national structure. The disadvantage is the indirect link between the product flow classifications.

In order to use physical flow volume data to estimate production, it is necessary to adjust the data for production flows within the regions (diagonal entries of the matrix), as well as for the effects caused by transport to and from warehouses. Production flows between the regions do not include these effects (they show net exports and imports between the regions without the effect of exports and imports of goods to and from warehouses). This effect systematically causes overvaluation of some regions of the Czech Republic, especially the Central Bohemian Region, Moravian-Silesian Region and Prague, where large warehouses are located. In order to adjust for this, it is necessary to reformulate the gravity model to produce the following model:

$$\tilde{x}_{i\bullet}^{pr} = G_i \frac{(x_{i\bullet}^{p\bullet})^{\alpha_i} (x_{i\bullet}^{\bullet r})^{\beta_i}}{\left(\delta^{pr}\right)^{\omega_i}} \left(l_p^{\psi_i} \big/ l_r^{\xi_i} \right), \tag{8}$$

for $r, p = 1,2......P$, (P=R), and $i, j = 1,2......n$.

Where $l_p^{\psi_j} l_r^{\xi_j}$ represents the number of warehouses in the region which exports (p) and imports (r), and where constants ψ_i, ξ_i represent the effect of this production flow to the warehouse for individual regions and products. However, the problem of this representation is the fact that the number of warehouses (or, more precisely, the overall warehouse space/area) is unknown at a certain amount. In order to adjust for this, we used the number of staff in the warehouses in approximation of the warehouse variable. This is followed by the standard procedure for estimating the gravity model – adjustment using the RAS method to meet the conditions of I-O tables, i.e.:

$$\hat{\mathbf{F}}_i = f_{RAS}(\tilde{\mathbf{F}}_i, x_{i\bullet}^{p\bullet}, x_{i\bullet}^{\bullet r}) \text{, for } r, p = 1,2......P, (P=R), \text{ and } i, j = 1,2......n. \tag{9}$$

[3] The NST classification is not proportional to the CZ-CPA classification, and this data only represents proxy structures for estimating the production flow matrices. Assignment of individual NSTs as a proxy variable of CZ-CPA is shown in Appendix No. 2.

4.2 Distance optimization (EO)

This approach is based on the assumption that the distance between regions plays a key role through costs. This means that individual regions try to minimize the import distance to minimize cost. Subsequently, this task can be formulated as follows:

$$\min_{x_{ij}^{pr}} f = \sum_{P=1}^{m}\sum_{i=1}^{m} \delta^{pr} x_{ij}^{pr} \tag{10}$$

With the limitations of the model:

$$\begin{aligned} x_{j\bullet}^{\bullet r} &= \sum\nolimits_{p}^{P}\sum\nolimits_{i}^{n} x_{ij}^{pr} \\ x_{i\bullet}^{p\bullet} &= \sum\nolimits_{r}^{R} x_{i\bullet}^{pr} \\ x_{i\bullet}^{\bullet r} &= \sum\nolimits_{p}^{P} x_{i\bullet}^{pr} \\ x_{ij}^{pr} &\geq 0 \\ x_{ij}^{rr} &= 0 \end{aligned} \tag{11}$$

The disadvantage of this approach is the fact that minimization of import distances leads to extreme results from the point of view of interconnection of links between the regions. This means that there are no links between some regions, but there are links between others (for individual products), which can even cause extreme results from the analytical point of view of data use. This has been confirmed in another application (Šafr, 2016a). Nevertheless, this approach has its own justification and use – for example, when assessing the impacts of infrastructure changes – Shortest Route Algorithm (Euijune, Hewings, 2009). However, compared to the previous approach, the model application results in direct estimates of matrix $\hat{\mathbf{F}}_i$, not of matrix $\widetilde{\mathbf{F}}_i$.

4.3 Calibration from physical flows (CPF)

This method is based on calibration of production flows according to the physical volume flow matrix between the regions in the NTS classification. The statistical yearbook of the Ministry of Transport can be used to obtain data on transport between individual regions of the Czech Republic in physical units. However, this approach has several problems. I) The data itself, which includes exports and imports to warehouses. II) It is necessary to use a matrix in another classification as a reference matrix for calibration of flow values. III) Physical volumes do not need to be directly related to the value representation of production flows in I-O data.

We dealt with the first problematic point using a regression estimate of this effect when, using an equation (the third at NGM), we estimated the effect of stock by region and product and, subsequently, we adjusted the data for this effect. The data adjustment effect can be seen in the Figure 1.

Such assigned flow matrices (adjusted, hereinafter referred to as $\widetilde{\Phi}_i$) will serve as the underlying structure for calibrating production flows between the regions. This can then be expressed as follows:

$$\hat{\mathbf{F}}_i = f_{RAS}(\widetilde{\Phi}_i, x_{i\bullet}^{p\bullet}, x_{i\bullet}^{\bullet r}) \tag{12}$$

for $r, p = 1,2......P$, (P=R), and $i, j = 1,2......n$

Thus, we obtain estimates of the production flow matrix. However, we must assume the relationship between physical volume and value representation in this estimation procedure, ideally direct linear.

4.4 Proportional distribution of flows (PDF)

With respect to the problematic character of production flow estimation, and because there is no "real" benchmark, we also used proportional distribution of production flows (Šafr, 2016a). This flow matrix estimate is, in essence, only application of the equation for unconditional probability from the contingency table, and can be expressed as follows:

$$\hat{x}_{i\bullet}^{pr} = \frac{x_{i\bullet}^{\bullet r} x_{i\bullet}^{p\bullet}}{\sum x_{i\bullet}^{p\bullet}} \tag{13}$$

for $r, p = 1,2......P$, (P=R), and $i, j = 1,2......n$

If export/import to regions is defined in balance[4] terms, this equation is the final estimate. If this relationship is not defined in balance terms, then it is necessary to set the diagonal entries of this matrix as equal to zero and to ensure that the assumptions are fulfilled by using the RAS method, as with the other methods here. With regard to its use, this model is essentially a partial decomposition of the gravity model in the case of equal distances between the regions, and the difference between the models then shows us the impact of distance and the coefficients of the model itself.

4.5 Inverse distance ratio (IDR)

This approach is based on the assumption that the flow volume between two regions is inversely proportional to their distance. Thus, this approach can be expressed as a RAS function:

$$\hat{\mathbf{F}}_i = f_{RAS}(\Delta^{\bullet -1}, x_{i\bullet}^{p\bullet}, x_{i\bullet}^{\bullet r}) \tag{14}$$

for $r, p = 1,2......P$, (P=R), and $i, j = 1,2......n$

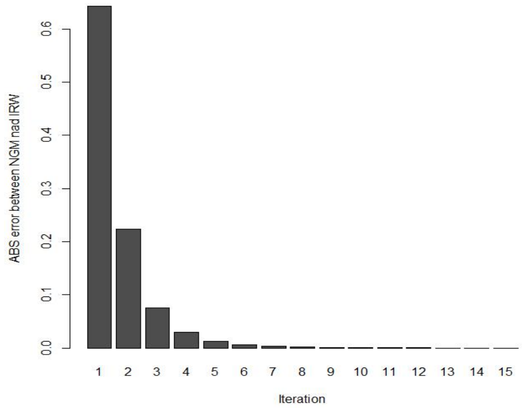
where $\Delta^{\bullet -1}$ is an inverse matrix (inversion by entries) of distances, $x_{i\bullet}^{p\bullet}$ is the exports of the product, and $x_{i\bullet}^{\bullet r}$ is the imports of the product – expressed as vectors whose elements are values for individual regions.

Although this approach may seem very "simple", even inappropriate, it can be shown that under the assumption of unit coefficients in the GM this approach has the same result as the GM. Due to the unavailability of data for calibration of coefficients, many authors (Kieslichová, 2016) often use this coefficient setting approach. Therefore, we assume that the GM often unknowingly leads to this approach. The following chart shows that with the increasing number of iterations the GM approximates the IDR, thanks to RAS iterative optimization.

Therefore, it can be said that the calculation of production flows through the matrix of inverse distances is a special case of the gravity model with unified model constants.

[4] Which means that, for a given product, a region can only be an importer/exporter. With regard to the Czech Republic data, tables have been produced for 2011 at the level of the CZ-CPA classification for 82 products. The balance representation itself will then provide zero diagonal entries in this calculation.

Figure 1: Convergence in the number of iterations of the RAS method



Source: The author's calculations

5. CGE model using regional data

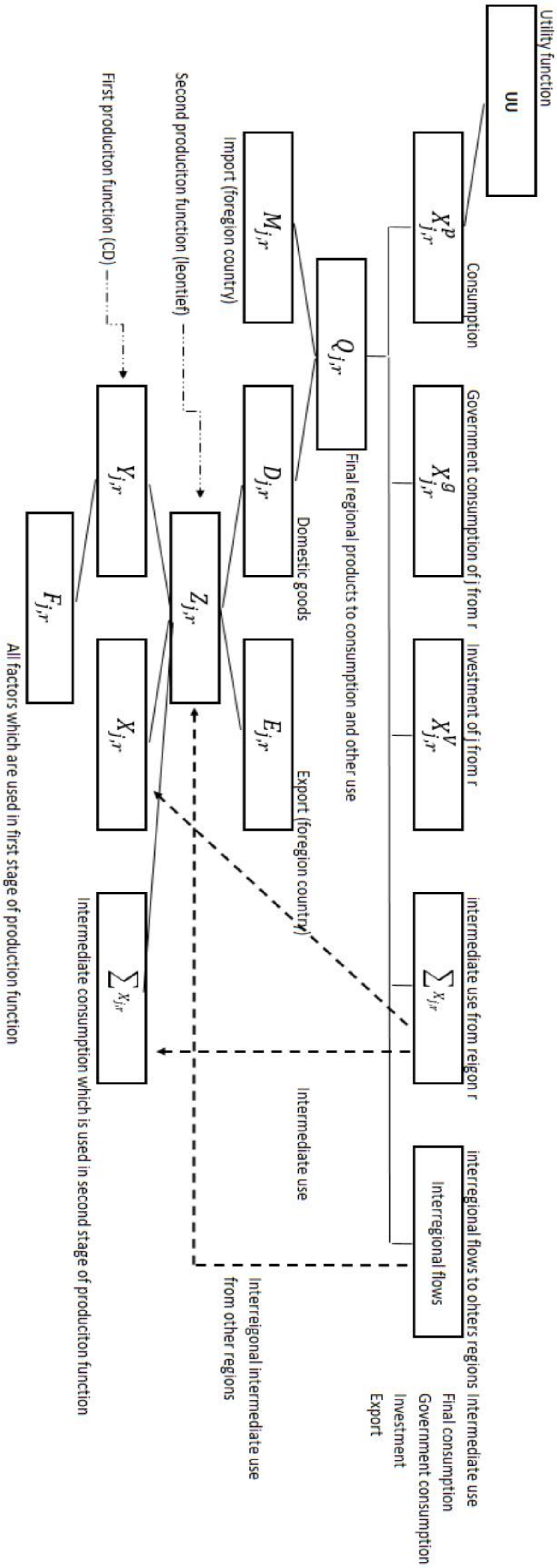
According to our approach, popular CGE models can also be constructed at the regional level. We use the standard concept of the model here. The main idea of the CGE model is the conflict between supply and demand, assuming general equilibrium. In our model, there are four actors in the regional distribution: consumers who maximize their benefits, companies that are owned by households and maximize their profits in the form of payments to households, the government that redistributes and levies taxes, and the foreign sector. This can be illustrated by the following diagram in Figure 2.

We have 8 aggregated regional products in the model (i.e. 64 products). All equations of our regional model are in the appendix (1). This model is constructed as disaggregation of the standard CGE model (Hosoe et al., 2010). The model uses regionally divided companies. Therefore, we differentiate in which region the product is produced. Thanks to this distinction, the consumer has a regionally structured consumer function, which allows us to make a smooth substitution between the products of individual regions. This is based on the estimation of regional production flows.

6. Data used

Our model integrates data of various sources, primarily information about the national accounts of the Czech Statistical Office (2017), from which we used export and import data in CIF/FOB valuation to estimate foreign import costs. In addition, there are regional Input-Output tables that were produced as part of a grant project at the Department of Economic Statistics of the University of Economics (KEST, 2017; Sixta, Vltavská, 2016). These regional tables then provide a basic framework for constructing estimates of interregional production flow matrices. Another important data source is the data from the statistical yearbooks of the Ministry of Transport (2017). This data served to estimate the impact of storage capacities on regional production flows. These regional production flows were then used in the GM and CPF data estimation approaches. As an additional data source, we used the Average Earnings Information System (2017), where employment in warehouses (codes) served as a proxy variable to adjust exports and imports for the impact of transport from and to warehouses. The last data source is the distances between individual regional centres, i.e. regional town. For this, we used the distance between the regional towns calculated as a lorry distance using Google's Maps data (Google, 2017).

Figure 2: Overview of the standard CGE model



Source: The author’s work based on standart CGE model by Hosoe et al. (2010)

7. Results obtained

Following the application of the aforementioned methods, we obtained 5 different intermediate consumption matrices for different model applications. These matrices are compared to each other in absolute values and also in the CGE model application, where we model the change in the tax rate in the price level and production volume.

7.1 Intermediate consumption matrix

Since the main goal was to construct production flows into intermediate consumption between individual regions, there is no difference in flow volumes as a whole (total sums of columns and rows do not change). The differences are in a different structure of the matrix itself. In order to assess the impact of various model applications, we calculated the absolute average from the difference of intermediate consumption matrices. The results (1) show that the greatest difference from other approaches is displayed by the linear optimization approach (EO – this was also confirmed in the article by Šafr (2016a). The gravity model (GM) and unconditional probability calculation (PDF) can be considered a medium approach. The greatest difference between the two approaches was between the unconditional probability method (PDF) and the linear optimization method (EO) – distance minimization. Conversely, the largest congruence occurred between the gravity model (GM) and the unconditional probability calculation method (PDF).

Table 1: Absolute average % distance between intermediate consumtion

	IDR	**EO**	**GM**	**CPF**	**PDF**	**MEAN**
CPF	0.00%	12.50%	4.44%	7.24%	4.52%	**5.74%**
EO	12.50%	0.00%	15.05%	12.96%	15.12%	**11.13%**
GM	4.44%	15.05%	0.00%	8.91%	0.38%	**5.76%**
IDR	7.24%	12.96%	8.91%	0.00%	8.94%	**7.61%**
PDF	4.52%	15.12%	0.38%	8.94%	0.00%	**5.79%**
MEAN	**5.74%**	**11.13%**	**5.76%**	**7.61%**	**5.79%**	

Source: The Author's calcuations

Differences from the gravity model are essential for us, because it is the most frequently used approach. It turned out that the results are very similar to the model where the calculation is estimated using unconditional probability (PDF) for both matrices. This confirms the assumption that the result in the gravity model is not primarily affected by distances, but by the absolute volumes of exports and imports of production for the regions in question. This is subsequently confirmed by the method based on the calculation from the inverse matrix of distances (IDR) – which is more distant from the gravity model than the PDF. As we have shown, the gravity method and the inverse distance calculation method are the same if the coefficients of the gravity model are not different from the unit. The difference between these two methods is the influence of the coefficients of the gravity model and, simultaneously, the degree of undervaluation/overvaluation that is incorporated in the model if the calculation of these coefficients is not considered (as in Kieslichová, 2016, etc.). Given that the unconditional probability calculation method represents a benchmark against which other methods can be defined, because it has very similar results to all other methods discussed, we plan to make more use of this method in the future.

7.2 Results of the CGE model

To assess the effect of these different regional matrices, we used a standard CGE model that we subjected to regional disaggregation. This model has all essential features of the CGE model for analytical use (i.e. government sector and foreign countries). Using this model, we modelled the tax burden increase (ad valorem taxes) in individual products, where we set the external shock value to 20%. In this assessment of the influence of a different data source on the mathematical application, we mainly focused on the differences of the results rather than on the results of the model, because the aim is to verify the quality of the regional structures. The following table summarizes the main results for individual products in the regions, expressing the % differences between the modelled impacts.

Table 2: % impact on production volume change:

	IDR	EO	GM	CPF	PDF	MEAN
IDR	0.00%	8.20%	2.62%	2.32%	2.56%	3.14%
EO	8.20%	0.00%	8.99%	8.64%	9.14%	6.99%
GM	2.62%	8.99%	0.00%	4.06%	0.91%	3.32%
CPF	2.32%	8.64%	4.06%	0.00%	3.79%	3.76%
PDF	2.56%	9.14%	0.91%	3.79%	0.00%	3.28%
MEAN	3.14%	6.99%	3.32%	3.76%	3.28%	

Source: The Author's calculations

The main differences are very similar to the differences in the entries calculated in the intermediate consumption matrix. The greatest difference in the methods modelled is the distance minimization method (EO). Conversely, the greatest congruence with other methods is achieved by the unconditional probability method (PDF). In this case, a greater congruence of the results (both for impacts on the price level and impacts on the overall production in the economy) was especially between the GM and PDF rather than the IDR.

The results show that in aggregation to the basic 8 products there is significant heterogeneity at the level of individual regions and products. The opposite effect, in terms of relatively consistent impacts, can be seen when we look at individual regions as a whole. Then, the average impact on production volumes and prices is relatively homogeneous. Nonetheless, dissimilar results can be found. Here, again, the most important are especially the differences from distance minimization (EO). The following table shows the % change in production in the region due to the 20% ad valorem tax.

Table 3: % impact on product price change:

	IDR	EO	GM	CPF	PDF	MEAN
IDR	0.00 %	4.64 %	1.19 %	4.64 %	1.57 %	2.41 %
EO	4.64 %	0.00 %	3.63 %	4.79 %	3.35 %	3.28 %
GM	1.19 %	3.63 %	0.00 %	1.16 %	0.56 %	1.31 %
CPF	4.64 %	4.79 %	1.16 %	0.00 %	1.51 %	2.42 %
PDF	1.57 %	3.35 %	0.56 %	1.51 %	0.00 %	1.40 %
MEAN	2.41 %	3.28 %	1.31 %	2.42 %	1.40 %	

Source: The Author's calculations

Table 4: % production change as a result of the shock of aggregate increase in taxes to 20% of the production value

REG|MET	IDR	EO	GM	CPF	PDF
CZ-NUTS 01	-5.41%	-6.19%	-4.52%	-5.77%	-4.28%
CZ-NUTS 02	-40.82%	-40.73%	-40.51%	-41.09%	-41.15%
CZ-NUTS 03	0.99%	3.55%	0.18%	1.22%	0.04%
CZ-NUTS 04	13.86 %	16.31 %	13.31 %	13.09 %	13.56 %
CZ-NUTS 05	-22.23%	-21.78%	-22.03%	-22.30%	-22.03%
CZ-NUTS 06	-22.14%	-21.17%	-23.02%	-22.35%	-22.89%
CZ-NUTS 07	26.13 %	20.13 %	26.40 %	26.30 %	26.39 %
CZ-NUTS 08	-19.31%	-19.01%	-20.04%	-18.71%	-20.02%
MEAN	-8.62%	-8.61%	-8.78%	-8.70%	-8.80%

Souce: The Author's calculations

8. Conclusion

Application of methods for estimating interregional flows shows that some methods have results more similar to each other than others. In particular, there is an evident significant relationship between the gravity model, the inverse distance optimization method and the estimation method based on unconditional probability. Conversely, the linear optimization method appears to be the most distant from all other methods. This is determined by the fact that it places emphasis on minimal trajectories, which is in line with our expectations. On the contrary, the method based on physical volumes of flows did not confirm the expected congruence with the gravity model. This difference can also be caused by many objective factors that distort this method. This can be due to high price heterogeneity or different price levels between the regions.

The quality of regionalization methods was assessed using a standard CGE model at the regional level. Given the price heterogeneity of the products/regions, the expected tax rate increase by 20% results in differences, especially at the level of individual products (but not regions as a whole). However, this is due to the restrictive conditions of the models for the estimation of production flows. It is known how much the region imports and exports, but it is not known where from and where to. However, it can be said that differences between the estimation methods are not negligible. In other words, if the CGE model formed the basis for further analysis (such as the CBA), these differences could have an impact on the rejection or confirmation of project implementation. In our view, a necessary deeper development of application of these methods is particularly important in expanding model applications. Here, these methods could be used, for example, for modelling the impacts of motorway completion or for modelling the impacts of regional protectionist measures on exports and imports from other regions. Significantly greater different impacts between the methods can then be expected for shocks that will only affect one industry or one region, or one industry in one region.

Acknowledgement

This paper has been prepared under the support of the project of the University of Economics, Prague - Internal Grant Agency, project No. 28/2017 “Regional statistics structures” and by Institutional Support for Long Period and Conceptual Development of Research and Science at Faculty of Informatics and Statistics, University of Economics, Prague.

References

[1] Average Earnings Information System. 2017. Archive 2013. TREXIMA, [cit. 10.2. 2017], https://www.ispv.cz/en/ispv-results/archive/2013.aspx.

[2] Cardenete, M. A., Sancho, F. 2004. Sensitivity of CGE Simulation Results to Competing SAM Updates, Review of Regional Studies, Vol. 34, No. 1, pp. 37–56.

[3] Czech Statistical Office. 2017. Databáze národních účtů. Praha: Český statistický úřad, 2017

[4] Euijune, K., Hewings, G. 2009. An Application of an integrated transport network – multiregional CGE model to the calibration of synergy effects of highway investments. Economic Systems Reserach. 2009, Vol. 21, No. 4. december, pp 377 – 397.

[5] Google, 2017, Google Maps direction for driving between czech regions, [cit. 16.2.2017] ,https://maps.google.com.

[6] Hosoe, N., Gasawa, K., Hashimoto, H., 2010. Textbook of Computable General Equillibrium Modeling. Palgrave Macmillan Uk. 235 pp. ISBN: 0230248144.

[7] KEST, 2017. Regionalization of gross domestic product employing expenditure approach (data). Katedra ekonomické statistiky, Fakulta informatiky a statistiky Vysoká škola ekonomická v Praze. [cit. 10. 2. 2017] http://kest.vse.cz/veda-a-vyzkum/vysledky-vedecke-cinnosti/regionalizace-odhadu-hrubeho-domaciho-produktu-vydajovou-metodou/.

[8] Kieslichová, K., 2016, Využití gravitačních modelů při konstrukci odhadů komoditních toků, Diploma Thesis, The University of Economics, Prague.

[9] Kim, E., Hewings, D. J., Hong, Ch., 2004. An Application of an Intergrated Transport Network-Multiregional CGE Model: A Framework for the Economic Analysis of Highway Projects. Economic Systems Research, Vol. 16, N. 3, pp. 235-258. ISSN: 1469-5758.

[10] Ministry of Transport. 2017. SYDOS – Dopravní statistika. [cit. 13.3.2017] www.sydos.cz

[11] Šafr, K. 2017. Kapitálové matice v regionální input-output analyse. Vědecká conference doktorského studia, fakulta informatiky a statistiky Vysoké školy ekonomické v Praze. 9. Únor 2017, pp. 173 – 181.

[12] Šafr, K., 2016a. Allocation of commodity flows in the regional Input-Output tables for the Czech Republic. In: Proceedings of 19th International Scientific Conference Application of Mathematics and Statistics in Economics. 31. srpna – 4. září 2016, Banská Štiavnica.

[13] Šafr, K., 2016b. Pilot Appliacation of the Dynamic Input-Output Model. Case Study of the Czech Republic 2005-2013. Statistika, 2016, Vol. 96, No. 2, pp. 15-31, ISSN 0322-788x.

[14] Sargento, A., L., M., Ramos, P., N., Hewings, G., J., D., 2012. Inter-regional Trade Flow Estimation through Non-survey Models: An Empirical Assessment. Economic Systems Research, Vol. 24, No. 2, pp. 173-1973, ISSN: 1469-5758.

[15] Shepherd, B., 2013. The Gravity Model of International Trade: A user Guide. United Nations Publicaiton. Thailand, Second ed. ISBN: 978-974-680-346-5.

[16] Sixta, J., Vltavská, K. 2016. Regional Input-output Tables: Practical Aspects of its Compilation for the Regions of the Czech Republic. Ekonomický časopis, Vol. 64, No. 1, pp. 56–69, ISSN 0013-3035.

Appendices

Appendix 1. Regional CGE MODEL based on standart CGE model (Hosoe et al, 2010)

Regional production:

$$Y_{j,p} = b_{j,p} \prod_h \left(F_{h,j,p}\right)^{\beta_{h,j,r,p}}, \qquad \forall i,p$$

$$F_{h,j,p} = \frac{\beta_{h,j,p} p^Y_{j,p}}{p^f_{h,p}} Y_{j,p}, \qquad \forall h,j,p$$

$$X_{i,j,r,p} = ax_{i,j,r,p} Z_{j,p}, \qquad \forall i,j,r,p$$

$$Y_{j,p} = ay_{j,p} Z_{j,p}, \qquad \forall j,p$$

$$p^z_{j,p} = ay_{j,p} p^y_{j,p} + \sum_r \sum_i ax_{i,j,r,p} p^q_{i,r} \qquad \forall j,p$$

Government:

$$T^d = \tau^d \sum_p \sum_h p^f_{h,p} FF_h$$

$$T^z_{j,p} = \tau^z_{j,p} p^z_{j,p} Z_{j,p} \qquad \forall j,p$$

$$T^m_{i,p} = \tau^m_{i,p} p^m_{i,p} M_{i,p} \qquad \forall i,p$$

$$X^g_{i,p} = \frac{\mu_{i,p}}{p^q_{i,p}} \left(T^d + \sum_p \sum_j T^Z_{j,p} + \sum_p \sum_i T^m_{i,p} - S^g \right) \qquad \forall i,p$$

Investment and savings:

$$X^v_{i,p} = \frac{\lambda_{i,p}}{p^q_{i,p}} \left(S^p + S^g + \varepsilon S^f \right) \qquad \forall i,p$$

$$S^p = ss^p \sum_p \sum_h p^f_{h,p} FF_h$$

$$S^g = ss^g \left(T^d + \sum_p \sum_j T^Z_{j,p} + \sum_p \sum_i T^m_{i,p} \right)$$

Regional households:

$$X^p_{i,p} = \frac{\alpha_{i,p}}{p^q_i} \left(\sum_p \sum_h p^f_{h,p} FF_h - S^p - T^d \right) \qquad \forall i,p$$

National export and imports:

$$p^e_{i,p} = \varepsilon p^{W_e}_{i,p} \qquad \forall i,p$$

$$p^m_{i,p} = \varepsilon p^{W_m}_{i,p} \qquad \forall i,p$$

$$\sum_p \sum_i p^{W_e}_{i,p} E_{i,p} + S^f = \sum_p \sum_i p^{W_m}_{i,p} M_{i,p}$$

Substitution between imports and domestic goods:

$$Q_{i,p} = \gamma_{i,p} \left(\delta m_{i,p} M^{\eta_{i,p}}_{i,p} + \delta d_{i,p} D^{\eta_{i,p}}_{i,p} \right)^{\frac{1}{\eta_{i,p}}} \qquad \forall i,p$$

$$M_{i,p} = \left[\frac{\gamma^{\eta_{i,p}}_{i,p} \delta m_{i,p} p^q_{i,p}}{\left(1 + \tau^m_{i,p}\right) p^m_{i,p}} \right]^{\frac{1}{1-\eta_i}} Q_{i,p} \qquad \forall i,p$$

$$D_{i,p} = \left[\frac{\gamma_{i,p}^{\eta_{i,p}} \delta d_{i,p} p_{i,p}^{q}}{p_{i,p}^{d}} \right]^{\frac{1}{1-\eta_i}} Q_{i,p} \qquad \forall i, p$$

Transformation between exports and domestic goods:

$$Z_{i,p} = \theta_{i,p} \left(\xi e_{i,p} E_{i,p}^{\phi_{i,p}} + \xi d_{i,p} D_{i,p}^{\phi_{i,p}} \right)^{\frac{1}{\phi_{i,p}}} \qquad \forall i, p$$

$$E_{i,p} = \left[\frac{\theta_{i,p}^{\phi_{i,p}} \xi e_{i,p} \left(1 + \tau_{i,p}^{z}\right) p_{i,p}^{z}}{p_{i}^{e}} \right]^{\frac{1}{1-\phi_{i,p}}} z_{i,p} \qquad \forall i, p$$

$$D_{i,p} = \left[\frac{\theta_{i,p}^{\phi_{i,p}} \xi d_{i,p} \left(1 + \tau_{i,p}^{z}\right) p_{i,p}^{z}}{p_{i}^{d}} \right]^{\frac{1}{1-\phi_{i,p}}} z_{i,p} \qquad \forall i, p$$

Market-clearing conditions:

$$Q_{i,p} = X_{i,p}^{q} + X_{i,p}^{g} + X_{i,p}^{v} + \sum_{j} X_{i,p} \qquad \forall i, p$$

$$\sum_{j} F_{h,j,p,} = FF_{h} \qquad \forall h$$

Variables

$Y_{j,p}$ - Composite factor produced in the first stage and used in the second stage by the *j*-th firm in the *p*-th region

$b_{j,p}$ - Scaling coefficient in the composite production function for the *j*-th firm in the p-th region.

$F_{h,j,p}$ - The *h*-th factor used in the *j*-th firm (at the first stage) in the *p*-th region.

$\beta_{h,j,r,p}$ - The input share coefficient in the composite factor production function of the *h*-th factor in the *j*-th firm (from the *r*-th to the *p*-th region)

$p_{j,p}^{Y}$ - The price of the *j*-th composite factor produced in the *p*-th reigon.

$p_{h,p}^{f}$ - The price of the *h*-th factor in the *p*-th region

$ax_{i,j,r,p}$ - The input-output coefficients for the goods from the *i*-th firm to the *j*-th firm from the *r*-th region to the *p*-th region.

$X_{i,j,r,p}$ - The intermediate input of the *i*-th firm (from the *r*-th reigon) used in the *j*-th firm (in the *p*-th region).

$Z_{j,p}$ - The gross domestic output of the *j*-th firm in the *p*-th region.

$ay_{j,p}$ - The input requirement coefficient of the *j*-th composite good in the *p*-th region for the output of the *j*-th good.

$p_{j,p}^{z}$ - Price of the gross domestic output of the *j*-th firm in the *p*-th region

$p_{i,r}^{q}$ - Price of the i-th composite good (the *r*-th reigon)

FF_{h} - Endowments of the h-th factor for the household

T^{d} - Sum of the direct tax

τ^{d} - The direct tax rate

$T_{j,p}^{z}$ - The production tax on the *j*-th good in the *p*-th region

$\tau_{j,p}^{z}$ - The production tax rate on the *j*-th good in the *p*-th reigon

$T_{i,p}^{m}$ - The import tariff on the *i*-th good in the *p*-th region

$p_{i,p}^{m}$ - The price of the import of the *i*-th good to the *p*-th region in domestic currency

$M_{i,p}$ -Import of the *i*-th good

$\tau_{i,p}^{m}$ - Import tariff rate of the *i*-th good in the *p*-th region

$X_{i,p}^{g}$ - Government consumption of the *i*-th good in the *p*-th region

$\mu_{i,p}$ -Government consumption share of the *i*-th good in the *p*-th region

$X_{i,p}^{v}$ - Investment of the *i*-th good in the *p*-th region

S^{p} - Household saving

S^{g} - Government saving

S^{f} -Foregin saving

ε - Foreign exchange rate (As ratio)

$\lambda_{i,p}$ - Investment share coefficient of the *i*-th good in the *p*-th region

ss^{p} -average propensity for savings by households

ss^{g} - average propensity for savings by government

$\alpha_{i,p}$ - coefficient of elasticity of consumption of the *i*-th goods in the *p*-th region

$p^{e}_{i,p}$ - price of export of the *i*-th good from the *p*-th region in domestic currency

$E_{i,p}$ -exports of the i-th good,

$\delta m_{i,p}, \delta d_{i,p}$ - input share coefficients in the Armington production function

$\phi_{i,p}$ - The elastiticity of substitution for the *i*-th good in the *p*-th region

$\eta_{i,p}$ - The parameter which defines the elasticity of substitution

$\gamma_{i,p}$ - The scaling coefficient (in the Arm. func.)

$\xi e_{i,p}$, $\xi d_{i,p}$ share coefficients for the i-th good transformation (for the *i*-th good in the *p*-th region)

$p^{W_e}_{i,p}$, $p^{W_m}_{i,p}$ price of the *i*-th good from the *p*-th region in foreign currency (export/import)

$Q_{i,p}$ - *i*-th Armingzon composite goods in the *p*-th region

$D_{i,p}$ - domestic goods of the *i*-th firm in the *p*-th region

Appendix 2. Used NTS classification for RAS to get output in CZ-CPA classification

NTS	CPA
NTS **01**	CZ-CPA **01–03**
NTS **02**	CZ-CPA **05–06**
NTS **03**	CZ-CPA **07–09**, **41–42**
NTS **04**	CZ-CPA **10–12**
NTS **05**	CZ-CPA **13–15**
NTS **06**	CZ-CPA **16–18**, **58–63**
NTS **07**	CZ-CPA **19**
NTS **08**	CZ-CPA **20–22**
NTS **09**	CZ-CPA **23**
NTS **10**	CZ-CPA **24–25**
NTS **11**	CZ-CPA **26–28**
NTS **12**	CZ-CPA **29–30**, **45–47**
NTS **13**	CZ-CPA **31–33**
NTS **14**	CZ-CPA **36–39**
NTS **15**	CZ-CPA **49–53**
NTS **18**	CZ-CPA **64–99**